

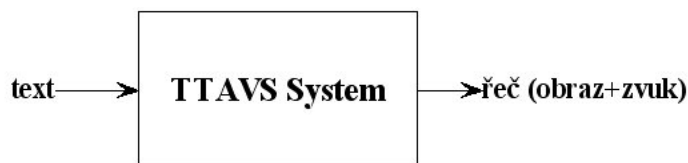


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

INTERNETOVÝ PORTÁL ELEKTROTECHNIKA - Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky
CZ.1.07/1.3.09/01.0021 D/0059/2009/ŘDP

Vizuální a akustická syntéza řeči

Syntézou řeči rozumíme napodobení lidské promluvy a v tomto případě využitím počítače. Tato problematika je obecně složena ze dvou částí: syntéza akustické řeči a syntéza vizuální řeči. Syntéza vizuální i akustické části řeči se často skrývá pod společným názvem "Mluvicí hlava" nebo také ve zkratce TTAVS systém. Syntéza akustické části řeči je vytváření právě té složky řeči, kterou můžeme slyšet a vizuální syntéza je pak tedy plynulé vytváření viditelné složky řeči, tedy to co můžeme pozorovat na tváři mluvícího člověka.

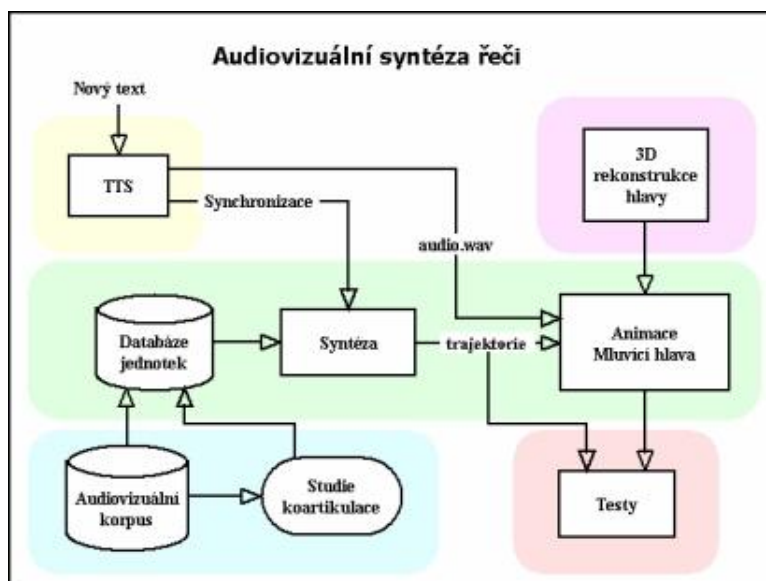


Syntéza vizuální řeči – animace tváře

Prostředek ke ztvárnění vizuální řeči může být nějaká animace lidské hlavy či alespoň tváře, nazývaná také "Mluvicí hlava".

Existuje několik přístupů řešení tohoto problému. Jsou navrženy syntézy používající klasický záznam promluvy lidské hlavy, syntéza je pak provedena spojováním vhodných částí těchto záznamů podle libovolného textu. Výhodou je veliká realističnost, nevýhodou pak velmi omezené ovládání animace, jako např. otočení hlavy, přiblížení, doplňování emocí apod. Dalším přístupem jsou 3D mluvící hlavy. Jde o pravé animace využívající nějaký model tváře, popř. model celé hlavy a očí, zubů, jazyka atd. Aby model mohl vytvářet vizuální řeč, musí být nějak řízen. Nejčastěji se setkáváme s tzv. parametrickým řízením. V tomto případě jde o to, že model je rozdělen na jednotlivé části, které podléhají určitému parametru. Hodnota tohoto parametru pak určuje míru změny (deformace) dané části. Např. můžeme uvést řízení pohybu levého koutku rtů, kdy jím můžeme posouvat ve všech třech směrech. Názor na to, jak parametrizace má vypadat však není ucelen. Existují řízení tváře velmi blízké anatomické podstatě tváře, modelování kostí, svalů a pokožky a naproti tomu více praktické

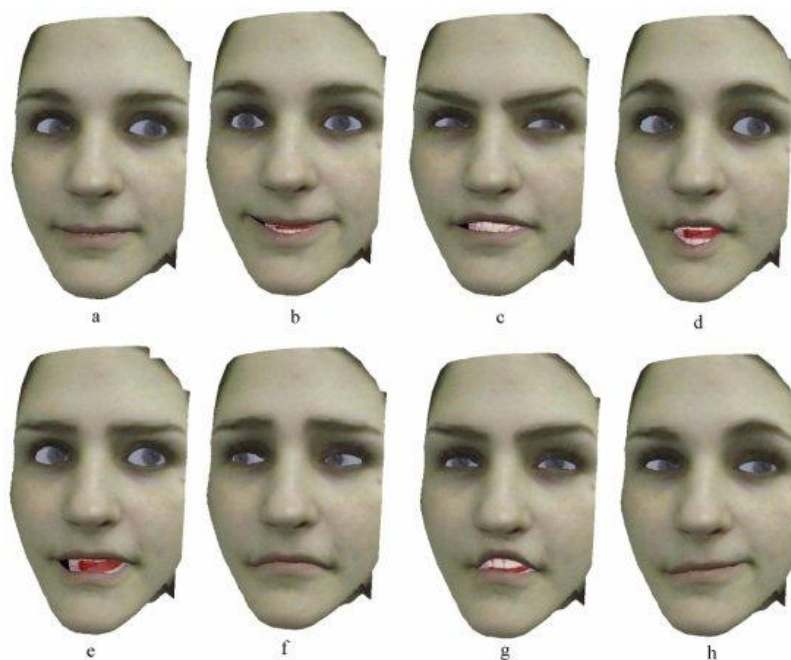
parametrizace, které jsou vhodně přizpůsobené animaci řeči. I známý standard kódování videa MPEG-4, popisující obecně kompresi audio-vizuálních dat, obsahuje parametrizaci tváře.



Záznam vizuální řeči

Problém záznamu vizuální řeči spočívá v přesném zachycení pohybů tváře. Nejvíce informativní je tvar rtů, dále pak pohyb tváří, brady a krku. Při řešení lze použít metodu sledování bodů pevně spojených s tváří. Pro tento účel se používají malé půlkulovité značky o průměru 3 a 4 mm. Kontura rtů je prozatím aproximována 8 body, dva body jsou na tvářích a po jednom na bradě a krku. Úkolem záznamu je učení pozice těchto značek v každém okamžiku promluvy. Pozice je dána 3D souřadnicemi a pro každých 20 či 40 ms máme jeden údaj. Záznam pohybů jazyka není v tomto návrhu možný. Jeho tvar, tj. šířka, poloha špičky a hřbetu je určena podle pravidel a doplněna ke každému vizému (skupina vzájemně si podobných hlásek). Jinou možností je však použití přesných zařízení jako ke snímkování magnetickou resonancí (MRI) nebo pomocí elektrod palatografu apod.

Animace výrazu tváře



a) neutrální, b) štěstí, c) rozzlobení, d) překvapení, e) strach, f) smutek, g) znechucení
h) zamyšlenost

Aplikace audiovizuální syntézy

Obecně jde o zařazení syntézy audio-vizuální řeči do procesu komunikace člověka s počítačem mluvenou řečí. Cílem je zlepšení porozumění řeči a to obzvláště na rušných místech jako je nádraží, letiště apod., ale i zlidštění a zpříjemnění dialogu včetně přenosu neverbální informací jako jsou např. různá emoce či gesta včetně znakové řeči. Výsledkem je systém, který poskytuje komplexní verbální i neverbální dialog. Aplikace audio-vizuální syntézy pak nacházíme především v informačních systémech, např. Mluvicí hlava "August" vyvíjená ve Švédsku. Systém je umístěný v centru Stockholmu a poskytuje potřebné turistické informace.